

Minireview

Wheat functional genomics and engineering crop improvementMichael Francki* and Rudi Appels*[†]

Addresses: *Department of Agriculture, Crop Improvement Institute, Locked Bag No 4, Bentley WA 6983, Australia. [†]CRC for Molecular Plant Breeding - Western node, School of Biological Science, Murdoch University, Murdoch WA 6150, Australia.

Correspondence: Michael Francki. E-mail: mfrancki@agric.wa.gov.au

Published: 23 April 2002

Genome **Biology** 2002, **3**(5):reviews1013.1–1013.5

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/5/reviews/1013>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

Abstract

Genetic mapping and determination of the organization of the wheat genome are changing the wheat-breeding process. New initiatives to analyze the expressed portion of the wheat genome and structural analysis of the genomes of *Arabidopsis* and rice are increasing our knowledge of the genes that are linked to key agronomically important traits.

The growing world-wide demand for wheat is placing pressure on breeding programs to produce elite cultivars that can adapt to a range of environments without compromising agronomic performance, grain quality or disease resistance. Wheat-breeding efforts focus on developing new varieties with improved disease resistance (to nematode, fungal and/or viral infection), tolerance to abiotic stresses (such as heavy-metal toxicity, drought and cold tolerance) and numerous grain quality attributes that affect baking and other uses of the final product. The combination of existing knowledge and resources with modern biotechnology and functional genomics is providing the opportunity to study the genetic, biochemical and physiological basis of these complex traits. Current efforts aim to address the major challenge of capturing the information from both wheat and model organisms, such as rice and *Arabidopsis*, in order to define genes that underpin the unique attributes of wheat. The resources being developed using biotechnology include comprehensive mapping initiatives and genome-wide expression studies; these need to be implemented together with wheat-breeding programs, in conjunction with high-throughput screening, to efficiently develop new, improved wheat varieties.

Wheat chromosome evolution

Bread wheat (*Triticum aestivum* L.) is an allohexaploid species containing three different ancestral genomes (designated A, B, and D), each of which has seven pairs of chromosomes

($2n = 6x = 42$). The total wheat genome is in excess of 16,000 megabases in size [1], and approximately 80% of it consists of repetitive (non-coding) DNA sequences, interspersed amongst low-copy-number or singleton genes. It is undoubtedly one of the largest and most complex genomes of all crop species.

The wheat genome can tolerate major chromosomal rearrangements and deletions [2,3] without major detrimental effects on the plant. Polyploidy provided a basis for chromosome engineering experiments [4] even before the cloning of DNA and the successful generation of transgenic plants. Significant improvements in wheat have been made by introgressing (transfer by hybridization and repeated back-crossing) chromosome segments from related grass species and the subsequent transfer of desirable genes. The most common introgression involves the short arm of rye chromosome 1 (1RS) translocated onto the long arm of wheat chromosome 1B (1BL), so that several disease-resistance genes from rye are transferred to wheat [5,6]. This whole-arm translocation is caused by a rare centromeric mis-division between wheat and rye univalent homeologous (closely related but not homologous) 1B chromosomes. The authors of recent studies have argued that combining genomes from different species can lead to chromosomal instability and rapid genome evolution [7], but the results of introducing chromosomes from related grasses into wheat have shown that many features of the introduced DNA remain stable. For example, we have analyzed a

number of wheat lines harboring 1BL.1RS translocations that show no changes in 1BL.1RS centric-fusion breakpoints [8] or other chromosomal regions (M.F. and W. Berzonsky, unpublished observations).

Genome sequence analyses have shed further light on the structural modifications that have occurred during the evolution of the wheat genome, particularly in intergenic regions. Alignment of a single four-gene region of *Triticum monococcum* (closely related to the wheat A genome), barley, and rice [9-11] indicated that gene order was maintained. Small inversions or duplications could occur, as indicated by gene 2 in this comparison, which was inverted in barley relative to its orientation in *T. monococcum* and rice, and by gene 4, which was duplicated in barley. The intron-exon structure of the predicted genes was conserved and could even be recognized in the distantly related *Arabidopsis* genome; the intergenic regions were quite divergent, however, as a result of the presence of different retrotransposable elements. In many instances the retrotransposable elements were nested, as described by San Miguel *et al.* [10], and estimates for the time of insertion, based on sequence divergence in the long terminal repeats of these elements, were less than 5 million years. The activity of some retrotransposable elements in rice as a result of cell culture [12,13] is consistent with the possibility that these elements provide an important means of genome modification.

Linking phenotypic traits to genes

Genetic mapping of wheat has provided the basis for identifying quantitative trait loci (QTL) that control complex traits [14]. The International Triticeae Mapping Initiative (ITMI) [15] has overseen the coordination of resources from different organizations world-wide to provide genetic mapping information. Langridge *et al.* [16] recently reviewed the development of a variety of genetic maps of wheat that serve a purpose for a number of applications: restriction fragment length polymorphisms (RFLPs), amplified fragment length polymorphisms (AFLPs) and simple sequence repeats (SSRs) are the predominant marker systems used, and together these provide numerous markers linking phenotypes to genes distributed across the wheat genome. As part of a major initiative to evaluate the expressed portion of the wheat genome, the US National Science Foundation (NSF) has funded an ongoing project to develop and sequence in excess of 100,000 expressed sequence tags (ESTs) [17]. ESTs are partial sequences derived from a range of cDNA libraries, providing a broad representation of expressed sequences from different tissue types during normal development or after treatment with abiotic and biotic stresses. To identify expressed sequences controlling specific traits, genetic mapping using ESTs as RFLP markers is hampered by the low-level polymorphism within wheat cultivars - polymorphism being a prerequisite for genetic-mapping studies. Well-characterized deleted regions of wheat chromosomes

have therefore been used to determine the physical position of all ESTs and to assign them to specific chromosome regions. This leads to a correlation of genetic and physical maps that will create a reference by which ESTs can be linked to QTLs. The NSF initiative is targeting 10,000 selected ESTs to specific chromosomal regions using deletion stocks of wheat [18].

Analyses of wheat ESTs from the mid-endosperm stage of seed development by our group and others [19-21] have revealed new predicted classes of seed storage protein that are most closely related to genes coding for avenin seed storage proteins from oat and foam-stabilizing proteins from barley. It is most likely that these new classes of gene encode proteins in the low-molecular-weight class of prolamin seed storage proteins described in earlier literature (discussed in [20,21]). Although this family of genes is extensive, we have localized some representatives to the long arm of chromosome 4A [20] using deletion stocks of wheat, and this chromosomal location is consistent with that of some low-molecular-weight prolamins determined by earlier researchers [22,23]. The wheat genes encoding foam-stabilizing-like proteins have been expressed in bacteria and the properties of the protein recovered in this way indicate they have significant influences on wheat-flour processing properties, an important functional attribute of seed storage proteins (B.C. Clarke and F. Bekes, personal communication).

Model systems for identifying candidate genes controlling complex traits

The completion of the *Arabidopsis* genome sequence [24] provided valuable information for the analysis of genome evolution across higher plants. For example, wheat starch properties are important in industrial and commercial uses, and specific starch-granule properties affect the quality of particular final products. The starch synthase III gene is involved in the production of the amylopectin class of starch, and the ratio of amylose to amylopectin has an impact on uses of the end product. A comparative analysis of the starch synthase III genes in wheat and *Arabidopsis* show that the exons are largely conserved, with some insertion events in wheat having introduced two additional exons and an enlarged exon 3 [25]. The large degree of conservation indicates that isoforms of starch synthase III have specific roles to play in starch synthesis across higher plants and that, in some instances, *Arabidopsis* genes can be used to identify corresponding genes in wheat.

Comparative genomics is contributing substantially to the analysis of genetically complex crops such as wheat. Although the much-discussed colinearity of grass genomes [26] is an over-simplification [27], because numerous inversion and insertion/deletion events have occurred during evolution, the concept of colinearity has provided a useful framework for analysis of the wheat genome. Many aspects

of the genome, such as intergenic regions and introns, have changed in sequence since wheat and other cereals diverged from their progenitor species, while the genic regions have been more conserved. Useful synteny between rice and wheat has been uncovered for several genes and for small regions of the genome [27], an important example being in the cloning of the *Rht-1* genes of wheat. The introduction of *Rht-1* genes into modern wheats reduced the plant's ability to respond to gibberellic acid (GA) [28] and allowed the selection of semi-dwarf wheat lines with much higher yields. The *Rht1* genes are functionally equivalent to the *Arabidopsis GA1* gene, which encodes a transcription factor integral to the plant's response to GA. *GA1* carries a characteristic phosphotyrosine signaling domain, and a rice EST with a high degree of similarity at the amino terminus was identified by Peng *et al.* [29]. The rice EST was subsequently used to identify the orthologous genes in wheat [29], which are located on the short arms of chromosomes 4B and 4D (near the centromere).

The *Rht-1* genes differ from the wild-type alleles found in most modern wheat by a single base-pair change that creates a TAG stop codon near the translation start site [29]. The polyploid nature of wheat means that each of the three genomes carries closely related sequences, however. In the case of the *Rht-1* genes, the mutants *Rht-B1b* (on chromosome 4B) and *Rht-D1d* (on chromosome 4D) clearly have to

be assayed individually. Differentiating between these closely related sequences is typical of the challenges faced in working with wheat, and Ellis *et al.* [30] were able to utilize sequence differences between the B and D genomes in the promoter regions of the genes to develop specific assays for the respective *Rht* genes. Figure 1 illustrates the location of the primers that were optimized to achieve the specific assay of the *Rht-B1b* and *Rht-D1d* mutant genes.

Engineering wheat improvement

Many marker-trait combinations are currently being implemented and tested in wheat-breeding programs. Advances in developing gene-specific markers from EST databases, with model systems, and with conventional DNA markers, in particular, are currently leading to strategies for identifying novel alleles in landraces or wild species that can be introgressed to improve wheat varieties. New alleles of complex traits can be identified in wild species using molecular markers [31] and then introduced into elite cultivars. Applying such strategies will broaden the gene pool for wheat improvement and maximize the genetic gain of desirable alleles from the parents in a common background during marker-assisted breeding. The relationship between increasing the number of parent alleles and maximizing genetic gain in a conventional and marker-assisted breeding is shown in Figure 2. Selection based on phenotyping is a major limitation

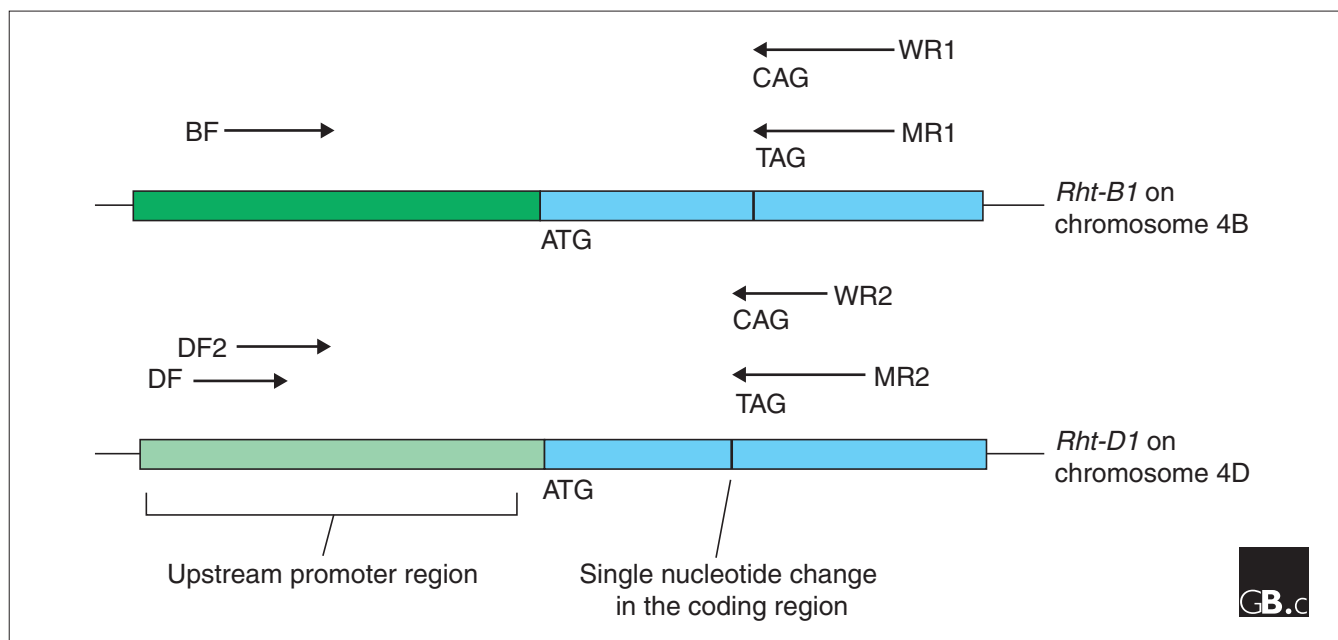


Figure 1

Diagram of the PCR primers used to distinguish the *Rht-B1b* and *Rht-D1b* genes [30]. The primers on the right differentiated the wild-type (WR1 and WR2) from the mutant (MR1 and MR2) genes by a single nucleotide change in the coding region. The primers cannot, however, distinguish between the genes on homeologous chromosomes 4B and 4D. The primers on the left-hand side differentiated between the 4B and 4D genomes because the upstream promoter regions have undergone sufficient evolutionary change to be distinguishable from each other. The figure is based on Ellis *et al.* [30] emphasizing the challenges faced when studying multiple gene copies in the wheat genome.

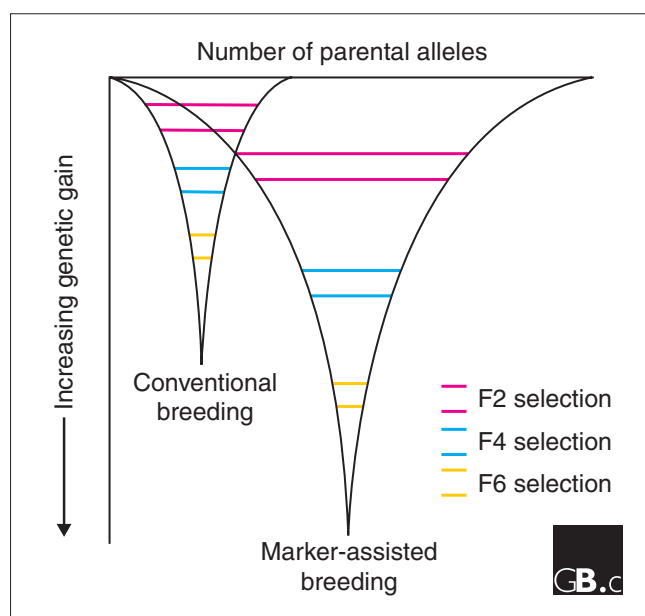


Figure 2

The effect on the genetic gain in elite lines of conventional and marker-assisted breeding. The effect of increasing the number of parental allele combinations, with selection at F2, F4 and F6 generations, highlights the value of a strategic use of molecular markers in breeding [32,33]. The larger base of the inverted triangle implies that more genetic variation is sampled for inclusion into the smaller numbers of lines examined in the F6 generation.

of screening large allelic combinations in a conventional breeding program, so increases in genetic gain are limited. Additional aspects of breeding to which molecular markers can contribute include recurrent parent selection in a back-crossing program, selection of parents for crossing, validation of F1s to ensure crosses have been made and exploitation of primitive germplasm with reference to reducing linkage drag by removing all unwanted chromosome segments more efficiently.

High-throughput screening capabilities now offer us the opportunity to select larger allelic combinations from wild and domesticated sources during marker-assisted breeding of wheat. Selecting progeny on the basis of genotyping at different stages of the breeding program will result in the accumulation of individuals with desirable allelic combinations and ultimately with a larger combination of desirable alleles at the culmination of the breeding program. Thus, understanding the complexities of the wheat genome leads to optimization of the technologies for combining desirable traits for improved wheat production.

References

1. Arumuganathan K, Earle ED: **Nuclear DNA content of some important plant species.** *Plant Mol Biol Rep* 1991, **9**:208-218.
2. Endo TR, Gill BS: **The deletion stocks of common wheat.** *J Hered* 1996, **87**:295-307.

3. Sears ER: **Nullisomic-tetrasomic combinations in hexaploid wheat.** In: *Chromosome Manipulation and Plant Genetics*. Edited by Lewis DR. London: Oliver and Boyd; 29-47.
4. Sears ER: **Chromosome engineering in wheat.** *Stadler Symp* 1972, **4**:23-38.
5. Baum M, Appels R: **The cytogenetic and molecular architecture of chromosome 1R - one of the most widely utilized sources of alien chromatin in wheat varieties.** *Chromosoma* 1991, **101**:1-10.
6. Berzonsky WA, Francki MG: **Biochemical, molecular and cytogenetic technologies for characterizing IRS in wheat: a review.** *Euphytica* 1999, **108**:1-19.
7. Ozkan H, Levy AA, Feldman M: **Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group.** *Plant Cell* 2001, **13**:1735-1747.
8. Francki MG, Berzonsky WA, Ohm HW, Anderson JM: **Physical location of a HSP70 gene homologue on the centromere of chromosome 1B of wheat (*Triticum aestivum* L.).** *Theor Appl Genet* 2002, **104**:184-191.
9. Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan L, Shiloff BA, Bennetzen JL: **Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes.** *Plant Physiol* 2001, **125**:1342-1353.
10. San Miguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards K, Lee M, Avramova Z, Bennetzen JL: **Nested retrotransposons in the intergenic regions of the maize genome.** *Science* 1996, **274**:765-768.
11. San Miguel PJ, Ramakrishna W, Bennetzen JL, Brusjo CS, Dubcovsky J: **Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5Am.** *Funct Integr Genomics* 2002, DOI 10.1007/s10142-002-0056-4.
12. Hirochika H: **Retrotransposons of rice as a tool for forward and reverse genetics.** In *Molecular Biology of Rice*. Edited by Shimamoto K. Tokyo: Springer; 1999: 43-58.
13. Agrawal GK, Yamazaki M, Kobayashi M, Hirochika R, Miyao A, Hirochika H: **Screening of the rice viviparous mutants generated by endogenous retrotransposon *Tos17* insertion.** *Plant Physiol* 2001, **125**:1248-1257.
14. Appels R, Gustafson JP, O'Brien L: **Wheat breeding in the new century: applying molecular genetic analyses of key quality and agronomic traits.** *Aust J Agric Res* 2001, **52**:1043-1043.
15. **International Triticeae Mapping Initiative** [<http://www.scri.sari.ac.uk/ITMI/>]
16. Langridge P, Lagudah ES, Holton TA, Appels R, Sharp PJ, Chalmers KJ: **Trends in genetic and genome analyses in wheat: a review.** *Aust J Agric Res* 2001, **52**:1043-1077.
17. **U.S. wheat genome project** [<http://wheat.pw.usda.gov/NSF/>]
18. **GrainGenes: Expressed Sequence Tags (GG-ESTs)** [<http://wheat.pw.usda.gov/wEST/binmaps/>]
19. Clarke BC, Hobbs M, Skylas D, Appels R: **Genes active in developing wheat endosperm.** *Funct Integr Genom* 2000, **1**:44-55.
20. Clarke BC, Larroque OR, Bekes F, Somers D, Appels R: **The frequent classes of expressed genes in wheat endosperm tissue as possible sources of genetic markers.** *Aust J Agric Res* 2002, **53**:1181-1193.
21. Anderson OD, Hsia CC, Adalsteins AE, Lew EJ-L, Kasarda DD: **Identification of several new classes of low molecular weight wheat gliadin related proteins and genes.** *Theor Appl Genet* 2001, **103**:307-315.
22. Salcedo G, Prada J, Aragoncillo C, Low MW: **Gliadin-like proteins from wheat endosperm.** *Phytochemistry*, 1979 **18**:725-727.
23. Rocher A, Calero M, Soriano F, Mendez E: **Identification of major rye secalins as coeliac immunoreactive proteins.** *Biochim Biophys Acta* 1996, **1295**:13-22.
24. **The Arabidopsis information resource** [<http://www.arabidopsis.org/>]
25. Li Z, Mouille G, Kosar-Hashemi B, Rahmen S, Clarke B, Gale KR, Appels R, Morrell MK: **The structure and expression of the wheat starch synthase III gene. Motifs in the expressed gene define the lineage of the starch synthase III gene family.** *Plant Physiol* 2000, **123**:613-624.
26. Devos KM, Gale MD: **Genome relationships: the grass model in current research.** *Plant Cell* 2000, **12**:637-646.
27. Lagudah ES, Dubcovsky J, Powell VV: **Wheat genomics.** *Plant Physiol Biochem* 2001, **39**:3335-344.
28. Gale MD, Gregory RS: **A rapid method for early generation selection of dwarf genotypes in wheat.** *Euphytica* 1977, **26**:733-738.

29. Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flinham JE, Beales J, Fish LJ, Worland AJ, Pelica F, *et al.*: **'Green revolution' genes encode mutant gibberellin response modulators.** *Nature* 1999, **400**:256-261.
30. Ellis MH, Spielmeier W, Gale KR, Rebetzke GJ, Richards RA: **'Perfect' markers for the Rht-B1b and Rht1-D1b dwarfing genes in wheat.** *Theor Appl Genet*, in press.
31. Xiao J, Grandillo LJ, Ahn SN, Yuan L, Tanksley SD, McCouch SR: **Identification of trait-improving quantitative trait loci from a wild rice relative, *Oryza rufipogon*.** *Genetics* 1998, **150**:899-909.
32. Ribaut J-M, Hoisington D: **Marker-assisted selection: new tools and strategy.** *Trends Plant Sci* 1998, **3**:236-239.
33. Barr AR, Jefferies SP, Warner P, Moody DB, Chalmers KJ, Langridge P: **Marker-assisted selection theory and practice.** In *Barley Genetics VIII - Proceedings of the 8th International Barley Symposium*. Edited by Logue S. Adelaide, Australia: Dept Plant Science, Adelaide University; 2000: 167-178.